DOWNLOADING AND USING DATA FROM THE STORET WAREHOUSE: AN EXERCISE

Sections 1 and 2 of this document walk you through the steps to generate and download a typical query from the STORET Data Warehouse. This query searches for data for specific characteristics found in a given 4th Code HUC. The 3rd section describes how to import the data into Excel to begin sorting and analyzing. The final section shows how to download data by watershed using the new Watershed Summary Tool. These exercises are intended to be used as a tutorial to help you become familiar with the STORET Warehouse. Please contact Jolene McQuillan at (406) 444-5304 or jmcquillan@mt.gov for questions or comments.

1   HOW TO QUERY AND DOWNLOAD DATA

For this example, our pre-planning has determined we need to search the STORET Warehouse for mercury, cadmium, and hardness results found in the Upper Missouri HUC, 10030101. The download file will be a compressed (zipped) .tar.gz file containing result data files, metadata files, and (if any) reference documents associated with the data owning organizations.

1.1   Go to the STORET Data Warehouse

1) Go to STORET Main Page: www.epa.gov/storet.
2) Under Features, click the **Download Data** link.
3) Under Modernized STORET, click the **Browse or Download Modernized STORET Data** button.

1.2   Define Your Query

4) Under Query Options, click the **Regular Results by Geographic Location** link.
   a) TIP: Queries by Geographic Location are best to obtain all available results. They search all organizations, projects, and stations. Queries by Project or Station are best to obtain results when searching for a known project or station(s).
5) Under Geographic Location, select the **Drainage Basin/HUC** radio button.
6) Under Cataloging Unit, click the **Look Up** button. Select "**10030101**" from the list and click the **Select** button.
7) Under Station Type, leave the defaults. This will capture all station types.
8) Under Date, leave the defaults. This will capture all date ranges.
9) Under Activity Medium, leave the defaults. This will capture all mediums.
10) Under Characteristic, place the cursor in the Characteristic Search box and type "**Mercury**". Click the **Search** button.
   a) TIP: Search box uses "Beginning with" type search (leading characters: "mer" works, not "cury").
   b) TIP: The percent sign, "%" is a wildcard search prefix that searches for parts of a characteristic name if the full name is unknown (example: typing "%cury" or "%cur%" all work).

11) Select "**Mercury**" from the Characteristic Name list, and click the **Select** button.
      a)   TIP: To select more than one Characteristic, use the Ctrl Key.
12) Search and select "**Cadmium**", "**Hardness, Ca + Mg**", "**Hardness, Calcium**", "**Hardness, Mg**", "**Calcium**", and "**Magnesium**".
13) Click the **Continue** button at the bottom of the screen.

1.3   Download Your Query Results

14) Note the number of records found.
      a)   TIP: Excel holds 65,000 records per sheet.
      b)   TIP: If there are too many records, you may have to go back and narrow your search by date or medium.
15) Scroll to the bottom of the screen and click the **Select All** button in the Select Data Elements for Report box.
      a)   TIP: These are the fields you will see in your report. Select all fields to make sure data is not being lost. Use Excel or Access to narrow down unwanted data after you download it.
16) Type your email address in the "Please enter your email address" box.
      a)   TIP: Emails are used to notify you that your download is first processing and then completed.
17) Type a three character prefix like "MER" in the "Please specify three characters to prefix your report name" box.
18) Under Batch Processing, click the **Immediate** button.
      a)   TIP: Immediate and Overnight reports follow the same directions, but small (<300K records) Immediate reports are available in 1-15 minutes while Overnight reports (600K max) are next day.
19) Go to your email account. You are waiting for two separate emails, a PROCESSING and COMPLETED email.
      a)   TIP: The COMPLETED email will contain a link to your file. The PROCESSING email can be deleted once the COMPLETED email has been received.
20) When you receive the COMPLETED email, your file is ready to download. Click on the **URL** in the email.
21) Click the **Save** button. Save the download file to your desktop or other directory and click **Close.**

2 MAKING SENSE OF YOUR DOWNLOADED FILE

Now that you have your downloaded file, what is it and what do you do with it? This section explains how to remove the data in your downloaded file and how to understand the file naming convention.

2.1 Retrieve Your Results Text File from the Download

1) Navigate to the directory where you saved the downloaded file from the STORET Data Warehouse.
2) Double-click the downloaded file to open it and click the **Yes** button when asked to decompress the file.
   a) TIP: Most compression engines like WINZIP will be able to open the .tar.gz file.
3) Extract all the files to a location on your computer for later access.

2.2 File Name Conventions

The files found in the download have four main components in their field name.

1) **Type of Document__:** Prefix denoting the document file is a data or reference document (Data__, RefDoc__).
2) **Unique Identifier__:** 3 char ID given, followed by the date/time stamp. (_XYZ_'year*mmdd*'_'24hr*mmss*'_).
3) **Type of Data__:** Suffix denoted the document contains results data, metadata, or reference data (_RegResults, _BioResults, _HabResults, _Metadata, _Project_'PROJECTID'_"filename").
4) **Type of File__:** Extension denoting the format of the document (.txt, .pdf, .bmp, .gif, .jpg).
5) Examples:
   a) Data_MER_20070428_140514_RegResults.txt
   b) Data_MER_20070428_140514_Metadata.txt
   c) RefDoc_MER_20070428_140514_Project_SWM_photo1.pdf

3 HOW TO IMPORT AND ORGANIZE THE DATA IN EXCEL

Now we're going to import the downloaded data into Excel and perform some basic operations to organize the data.

3.1 Import the Data into Excel

1) Open Excel.
2) From the main toolbar, click **Data→Import External Data→Import Data**.
3) In the Select Data Source window, navigate to the directory or folder where you have saved your downloaded STORET file.
4) In File of Type textbox, select "**All Data Sources**" or "**Text File**" from the drop-down list.
5) Select and double click your saved STORET File.
6) In the Text Import Wizard, click the **Delimited** radio button and then click **Next**.
7) Uncheck the **Tab** box and check the **Other** box. Type the tilde character "**~**" into the field next to Other.
8) Click **Next** and then click the **Finish** button.
9) Click the **Existing Worksheet** radio button and then click **OK**.

3.2 Freeze Panes

10) Select cell A2.
11) From the main toolbar, click **Window→Freeze Panes**.
    a) TIP: With freeze panes, you can scroll up and down the spreadsheet and still see the column headings.

3.3 Filter the Data

12) From the main toolbar, click **Data→Filter→Auto Filter**.
    a) TIP: The column headings will now have drop-down menus that allow you to view the unique contents in each column.
    b) TIP: Use the filter feature to determine if columns have data. If they do not, delete the column to make the worksheet more manageable.

3.4 Sort the Data

13) Select the whole worksheet by clicking the gray box in the upper-left corner of the worksheet.
14) From the main toolbar, click **Data→Sort**.
15) Using the Sort by drop-down list, select **Activity Start** and leave the radio button next to Ascending selected.

4   DOWNLOADING A WATERSHED SUMMARY

This exercise walks you through the steps to answer the question, "What data is available for the Helena watershed?" Using the Watershed Summary Tool, we will generate and download a query for all characteristics found in a given Watershed from the STORET Data Warehouse.

4.1   Go to the STORET Data Warehouse and Watershed Summary

1) Go to STORET Main Page: www.epa.gov/storet.
2) Under Features, click the **Download Data** link.
3) Under Modernized STORET, click the **Browse or Download Modernized STORET Data** button.
4) Under STORET Warehouse Watershed Summary, click **Summary Results by Watershed** link.

4.2   Determine Your Watershed (Skip 5.2 if you know your HUC(s))

5) On the STORET Warehouse Watershed Summary page, click the **Surf Your Watershed** link then click the **Locate Your Watershed** icon.
6) Scroll to the bottom of the page, find the Locate by Geographic Unit box, and use the drop-down menu to select **City Name**.
7) In the Your Entry textbox, type **Helena** and click the **Submit** button.
8) Write down the HUC(s) from your search results to define your query later. For Helena, there is only one HUC, 10030101.
9) Return to the STORET Warehouse Watershed Summary page by using the **Back** button.

4.3   Define Your Query

10) On the STORET Warehouse Water Summary page, click the **Go to Watershed Summary** link.
11) Under Drainage Basin/HUC, select **10030101 – Upper Missouri** from the drop-down list.
    a) TIP: This is the information you wrote down from step 8.
12) Under Search By, click the **Characteristic Type** radio button and then click the **Query** button. The types of characteristics available in HUC 10030101 are displayed along with the number of records (results) for each characteristic.
13) Click the **Nutrient** link.
    a) Browse the characteristics to see which characteristics (like Phosphorus) fall into this category.
    b) Browse the records list to see how many records of each characteristic exist in this particular HUC.
14) Click **Close** and repeat step 13 above for any other category links that you are interested in.
15) Click the **Select All** button. This gives you all the data in HUC 10030101.
    a) TIP: You can narrow your query by checking only categories you care about.
16) Click the **Get Results** button.

a) TIP: You may need to turn off your internet browser's pop-up blocker to allow the results pop-up windows to open.

17) Use Sections 1.3, 2, and 3 to download your query results and import them into Excel.

18) If your watershed contains multiple HUCs, repeat steps 10-17 for the remaining HUCs.